What is Shannon information?

Olimpia Lombardi[1] – Federico Holik[2] – Leonardo Vanni[3]

[1] CONICET – Universidad de Buenos Aires

[2] CONICET – Universidad Nacional de la Plata

[3] Universidad de Buenos Aires

1.- Introduction

Although the use of the word 'information', with different meanings, can be traced back to antique and medieval texts (see Adriaans 2013), it is only in the 20th century that the term begins to acquire the present-day sense. Nevertheless, the pervasiveness of the notion of information both in our everyday life and in our scientific practice does not imply the agreement about the content of the concept. As Luciano Floridi (2010, 2011) stresses, it is a polysemantic concept associated with different phenomena, such as communication, knowledge, reference, meaning, truth, etc. In the second half of the 20th century, philosophy begins to direct its attention to this omnipresent but intricate concept in an effort of unravel the tangle of significances surrounding it.

According to a deeply rooted intuition, information is related with data, it has or carries content. In order to elucidate this idea, the philosophy of information has coined the concept of semantic information (Bar-Hillel and Carnap 1953, Bar-Hillel 1964, Floridi 2013), strongly related with notions such as reference, meaning and representation: semantic information has intentionality –"aboutness"–, it is directed to other things. On the other hand, in the field of science certain problems are expressed in terms of a notion of information amenable to quantification. At present, this mathematical perspective for understanding information is manifested in different formalisms, each corresponding to its own concept: Fisher information (which measures the dependence of a random variable X on an unknown parameter θ upon which the probability of X depends; see Fisher 1925), algorithmic information (which measures the length of the shortest program that produces a string on a universal Turing machine; see, e.g., Chaitin 1987), von Neumann entropy (which gives a measure of the quantum resources necessary to faithfully encode the state of the source-system; see Schumacher 1995), among others. Nevertheless, it is traditionally agreed that the seminal work for the mathematical view of information is the paper where Claude Shannon (1948) introduces a precise formalism designed to solve certain specific technological problems in communication engineering (see also Shannon and Weaver 1949). Roughly speaking, Shannon entropy is concerned with the statistical properties of a given system and the correlations between the states of two systems, independently of the meaning and any semantic content of those states. Nowadays, Shannon's theory is a basic ingredient of the communication engineers training.

At present, the philosophy of information has put on the table a number of open problems related with the concept of information (see Adriaans and van Benthem 2008): the possibility of unification of various theories of information, the question about a logic of information, the relations between information and thermodynamics, the meaning of quantum information, the links between information and computation, among others. In this wide panoply of open issues, it can be supposed that any question about the meaning and interpretation of Shannon information has a clear and undisputed answer. However, this is not the case. In this paper we will see that, in spite of the agreement concerning the traditional and well understood formalism, there are many points about Shannon's theory that still remain obscure or have not been sufficiently stressed. Moreover, the very interpretation of the concept of information is far from unanimous.

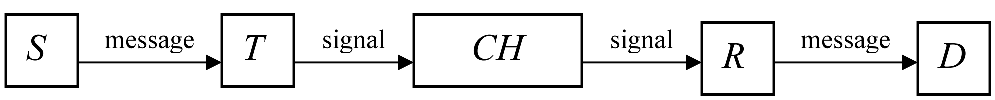
In order to develop the argumentation, Section 2 will begin by recalling the basic formalism of Shannon's theory. In Section 3, the abstract nature of information will be considered, and in Section 4 the fact that the theory deals with averages will be stressed. The gradual transformation experienced by the meaning of the term 'bit' during the last decades will be pointed out in Section 5. In Section 6, the relation between the definition of information and the coding of information will be discussed. The next two sections will be devoted to argue for the relative nature of information (Section 7), and for the theoretical neutrality of information (Section 8). The differences between two traditional interpretations of the concept information in the context of Shannon's theory, the epistemic and the physical interpretations, will be emphasized in Section 9. This task will allow us to propose, in Section 10, a formal reading of the concept of Shannon information, according to which the epistemic and the physical views are different possible models of the formalism.

2.- Shannon's theory

With his paper "The Mathematical Theory of Communication" (1948), Shannon offered precise results about the resources needed for optimal coding and for error-free communication. This paper was immediately followed by many works of application to fields as radio, television and telephony. Shannon's theory was later mathematically axiomatized (Khinchin 1957).

According to Shannon (1948; see also Shannon and Weaver 1949), a general communication system consists of five parts:

- A *source S*, which generates the message to be received at the destination.
- A *transmitter T*, which turns the message generated at the source into a signal to be transmitted. In the cases in which the information is encoded, encoding is also implemented by this system.
- A *channel CH*, that is, the medium used to transmit the signal from the transmitter to the receiver.
- A *receiver R*, which reconstructs the message from the signal.
- A *destination D*, which receives the message.



The source S is a system with a range of possible states $s_1,...,s_n$ usually called *letters*, whose respective probabilities of occurrence are $p(s_1),...,p(s_n)$.[1] The amount of information generated at the source by the occurrence of s_i is defined as:

$$I(s_i) = \log(1/p(s_i)) = -\log p(s_i) \tag{1}$$

Since S produces sequences of states, usually called *messages*, the *entropy of the source S* is defined as the average amount of information produced at the source:

$$H(S) = \sum_{i=1}^{n} p(s_i)\log(1/p(s_i)) = -\sum_{i=1}^{n} p(s_i)\log p(s_i) \tag{2}$$

Analogously, the destination D is a system with a range of possible states $d_1,...,d_m$, with respective probabilities $p(d_1),...,p(d_m)$. The amount of information $I(d_j)$ received at the destination by the occurrence of d_j is defined as:

$$I(d_j) = \log(1/p(d_j)) = -\log p(d_j) \tag{3}$$

[1] Here we are considering the discrete case, but all the definitions can be extended to the continuous case (see, e.g., Cover and Thomas 1991).

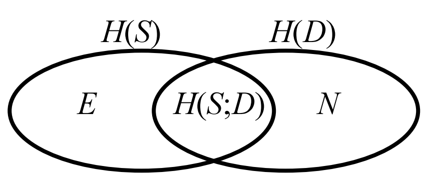
And the *entropy of the destination D* is defined as the average amount of information received at the destination:

$$H(D) = \sum_{j=1}^{m} p(d_j) \log(1/p(d_j)) = -\sum_{j=1}^{m} p(d_j) \log p(d_j) \tag{4}$$

In his original paper, Shannon (1948, p. 349) explains the convenience of the use of a logarithmic function in the definition of the entropies: it is practically useful because many important parameters in engineering vary linearly with the logarithm of the number of possibilities; it is intuitive because we use to measure magnitudes by linear comparison with unities of measurement; it is mathematically more suitable because many limiting operations in terms of the logarithm are simpler than in terms of the number of possibilities. In turn, the choice of a logarithmic base amounts to a choice of a unit for measuring information. If the base 2 is used, the resulting unit is called '*bit*' –a contraction of *binary unit* –. With these definitions, one bit is the amount of information obtained when one of two equally likely alternatives is specified.

It is quite clear that $H(S)$ and $H(D)$ are average amounts of information. Nevertheless, in the literature they are usually termed 'entropies', a strategy that could be explained by the fact that it is a name shorter than 'average amount of information'. However, according to a traditional story, the term 'entropy' was suggested by John von Neumann to Shannon in the following terms: "*You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name. In the second place, and more importantly, no one knows what entropy really is, so in a debate you will always have the advantage*." (Tribus and McIrving 1971, p. 180). In Italian it is usually said: "*se non è vero, è ben trovato*", that is, "*if it is not true, it is a good story*". In fact, even at present there are still many controversies about the content of the concept of entropy, whose deep implications can be easily compared to those resulting from the debates about the meaning of the term 'information'.

The relationship between the entropies of the source $H(S)$ and of the destination $H(D)$ can be represented in the following diagram:



where:

- $H(S;D)$ is the *mutual information*: the average amount of information generated at the source *S* and received at the destination *D*.
- E is the *equivocation*: the average amount of information generated at *S* but not received at *D*.
- N is the *noise*: the average amount of information received at *D* but not generated at *S*.

As the diagram clearly shows, the mutual information can be computed as:

$$H(S;D) = H(S) - E = H(D) - N \tag{5}$$

Equivocation *E* and noise *N* are measures of the dependence between the source *S* and the destination *D*:

- If *S* and *D* are completely independent, the values of *E* and *N* are maximum ($E = H(S)$ and $N = H(D)$), and the value of $H(S;D)$ is minimum ($H(S;D) = 0$).
- If the dependence between *S* and *D* is maximum, the values of *E* and *N* are minimum ($E = N = 0$), and the value of $H(S;D)$ is maximum ($H(S;D) = H(S) = H(D)$).

The values of *E* and *N* are functions not only of the source and the destination, but also of the communication channel *CH*. The introduction of the communication channel leads directly to the possibility of errors in the process of transmission: the channel *CH* is defined by the matrix $\left[p(d_j/s_i) \right]$, where $p(d_j/s_i)$ is the conditional probability of the occurrence of d_j in the destination *D* given that s_i occurred in the source *S*, and the elements in any row add up to 1. On this basis, *E* and *N* can be computed as:

$$N = \sum_{i=1}^{n} p(s_i) \sum_{j=1}^{m} p(d_j/s_i) \log(1/p(d_j/s_i)) \tag{6}$$

$$E = \sum_{j=1}^{m} p(d_j) \sum_{i=1}^{n} p(s_i/d_j) \log(1/p(s_i/d_j)) \tag{7}$$

where $p(s_i/d_j) = p(d_j/s_i)\,p(s_i)/p(d_j)$. The *channel capacity C* is defined as:

$$C = \max_{p(s_i)} H(S;D) \tag{8}$$

where the maximum is taken over all the possible distributions $p(s_i)$ at the source. C is the largest average amount of information that can be *transmitted* over the communication channel CH.

The two most important results obtained by Shannon are the theorems known as *First Shannon Theorem* and *Second Shannon Theorem*. According to the First Theorem, or *Noiseless-Channel Coding Theorem*, for sufficiently long messages, the value of the entropy $H(S)$ of the source is equal to the average number of symbols necessary to encode a letter of the source using an ideal code: $H(S)$ measures the optimal compression of the source messages. The proof of the theorem is based on the fact that the messages of N letters produced by S fall into two classes: one of approximately $2^{NH(S)}$ typical messages, and the other of atypical messages. When $N \to \infty$, the probability of an atypical message becomes negligible; so, the source can be conceived as producing only $2^{NH(S)}$ possible messages. This suggests a natural strategy for coding: each typical message is encoded by a binary sequence of length $NH(S)$, in general shorter than the length N of the original message.

On the other hand, in the early 1940s, it was thought that the increase of the rate in the information transmission over a communication channel would always increase the probability of error. The Second Theorem, or *Noisy-Channel Coding Theorem*, surprised the communication theory community by proving that that assumption was not true as long as the communication rate was maintained below the channel capacity. The channel capacity is equal to the maximum rate at which the information can be sent over the channel and recovered at the destination with a vanishingly low probability of error.

The formal simplicity of Shannon's theory might suggest that the interpretation of the involved concepts raises no difficulty. As we will see in the following sections, this is not the case at all.

3.- A quantitative theory without semantic dimension

One of the most cited quotes by Shannon is that referred to the independence of his theory with respect to semantic issues: "*Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages.*" (Shannon 1948, p. 379).

Many authors are convinced that the elucidation of a philosophically technical concept of semantic information, with his links with knowledge, meaning and reference, makes sense (see, e.g., Barwise and Seligman 1997, Floridi 2013). Moreover, there have been attempts to add a semantic dimension to a formal theory of information, in particular, to Shannon's theory (MacKay 1969; Nauta 1972; Dretske 1981). Although very fruitful, these approaches do not cancel the fact that Shannon's theory, taken by itself, is purely quantitative: it ignores any issue related to informational content. Shannon information is not a semantic item: semantic items, such as meaning, reference or representation, are not amenable of quantification. Therefore, the issue about possible links between semantic information and Shannon information is a question to be faced once the concept of Shannon information is endowed with a sufficiently clear interpretation. But precisely this is still the step to be done: what is the nature of Shannon information?

During the last years, in the philosophy of physics community (not in the physics community) it is usual to hear that the problem of the interpretation of the concept of Shannon information is dissolved because the word 'information' is an abstract noun. Christopher Timpson (2004, 2008, 2013) advocates for this position by stressing the analogy between 'truth' and 'information' and recalling Austin's claims about the notion of truth: "*Austin's aim was to de-mystify the concept of truth, and make it amenable to discussion, by pointing to the fact that 'truth' is an abstract noun. So too is 'information'.*" (Timpson 2004, p. 3). His best-known argument is based on the philosophical distinction between *types* and *tokens*: if the source produces the sequence of states $s_8, s_5, s_1, s_2, s_2, s_4, ..., s_7, s_7, s_2, ..., s_9, s_3, s_1$, what we want to transmit is not the sequence of states itself, but another token of the same type: "*one should distinguish between the concrete systems that the source outputs and the type that this output instantiates.*" (Timpson 2004, p. 22; see also Timpson 2008, 2013). The goal of communication is, then, to reproduce at the destination another token of the same type: "*What will be required at the end of the communication protocol is either that another token of this type actually be reproduced at a distant point*" (Timpson 2008, p. 25). Once this point is accepted, the argument runs easily: since the information produced by the source, that we desire to transmit, is the sequence type, not the token, and types are abstract, then information is abstract and 'information' is an abstract noun (see Timpson 2004, pp. 21-22; see also 2008).

Although seemingly convincing, some doubts soon arise: if the information is the type transmitted, and information can be measured, what is the measure of a type? Moreover, is it true that the goal of communication is to reproduce at the destination a token of the same type produced at the source? Timpson explains that "*if the source X produces a string of letters like the following:* $x_2, x_1, x_3, x_1, x_4, ..., x_2, x_1, x_7, x_1, x_4$ *, say, then the type is the sequence '*$x_2, x_1, x_3, x_1, x_4, ..., x_2, x_1, x_7, x_1, x_4$*'; we might name this 'sequence 17'. The aim is to produce at the receiving end of the communication channel another token of this type. What has been transmitted, though, the information transmitted on this run of the protocol, is sequence 17.*" (2004, pp. 21-22). But this is not the case: as Shannon stresses, in communication "[t]*he significant aspect is that the actual message is one* selected from a set *of possible messages*." (Shannon 1948, p. 379; emphasis in the original). This means that what has been transmitted is not sequence 17, but that sequence 17 is the actual message *selected from the set* of the possible messages of the source. As a consequence, the states d_j of the destination system *D* can be any kind of states, completely different from the states s_i of the source system *S*: the goal of communication is to identify at the destination which sequence of states s_i was produced by the source. Indeed, the destination can identify that sequence 17 was produced in the source by the occurrence in *D* of a sequence $d_7, d_4, d_3, d_4, d_5, ..., d_7, d_4, d_7, d_4, d_5$, which can hardly be regarded as a token of the type '$x_2, x_1, x_3, x_1, x_4, ..., x_2, x_1, x_7, x_1, x_4$'. Therefore, in principle the sequences of the source and of the destination do not need to be tokens of the same type in any sense that does not empty the very philosophical distinction type-token of any content.

In one of his papers about the notion of information, Armond Duwell notes that: "*To describe the success criterion of Shannon's theory as being the reproduction of the tokens produced at the information source at the destination is unacceptable because it lacks the precision required of a success criterion.*" (Duwell 2008, p. 199). The reasons are several: (i) any token is a token of many different types simultaneously; so the type-token argument leaves undetermined the supposedly transmitted type (*ibid*. p. 199), (ii) the Shannon entropy associated with a source can change due to the change of the probability distribution describing the source, without the change of the types that the source produces tokens of (*ibid*. p. 202), and (iii) the types a source produces tokens of can change without the Shannon entropy of the source changing (*ibid*. 203). But the main reason is that, in Shannon's theory, the success criterion is given by a one-one mapping from the set of letters that characterize the source to the set of letters that characterize the destination, and this mapping is completely *arbitrary* (*ibid*. p. 200).

Therefore, the states of the source and the states of the destination may be of a completely different nature: for instance, the source may be a dice and the destination a dash of lights; or the source may be a device that produces words in English and the destination a device that operates a machine. It is difficult to say in what sense a face of a dice and a light in a dash are tokens of a same type and which the type is in this case. The fact that any token is a token of different types does not mean that any two things arbitrarily chosen can always be conceived as tokens of the same type: admitting arbitrary functions as defining the relation "*x* is a token of the same type as the token *y*" deprives the distinction type-token of any philosophical content and conceptual usefulness (see Wetzel 2011).

With the purpose of retaining Timpson's proposal in spite of criticisms, Duwell recalls the distinction, introduced by Timpson (2004, pp. 20-21), between *Shannon quantity-information*, which "*is that which is quantified by the Shannon entropy*" (Duwell 2008, p. 201), and *Shannon type-information*, which "*is what is produced at the information source that is required to be reproduced at the destination*" (*ibid*., p. 201).[2] However, far from elucidating Shannon's concept of information, this distinction makes clear that the information usually measured in bits and which engineers are really interested in is the quantity-information, which is not a type and has nothing to do with types and tokens.

Summing up, by contrast with types and tokens, Shannon information is a measurable item. Although the idea of type-information does not imply to endow types with meaning, a type needs to have some content to be able to identify its tokens: the distinction between types and tokens is not merely syntactic. On the contrary, Shannon information is neutral with respect to any content, since the only relevant issue is the selection of a message among many.

4.- A theory about averages

In Section 2, the entropies of the source $H(S)$ and of the destination $H(D)$ were conceived as average amounts of information, defined in terms of the individual amounts of information $I(s_i)$ and $I(d_j)$, corresponding to the occurrence of a single state of the source and of the destination,

[2] In spite of criticisms, Duwell makes a strong argumentative effort to retain Timpson's proposal. For instance, he distinguishes between the *success* of communication and the *goal* of communication, which "*is to produce, at the destination, a token of the type produced by the information source.*"(Duwell 2008, p. 199). In this way, he retains Timpson's view at the cost of introducing notions, the goal of communication and Shannon type-information, which are completely absent in Shannon's theory.

respectively (eqs. (1) and (3)). However, in some presentations of Shannon's theory the individual amounts of information do not even appear, and the entropies $H(S)$ and $H(D)$ are defined directly in terms of the probabilities of the states of the source and the destination according to eqs. (2) and (4), respectively.

On this basis, some authors never talk about the information generated or received by a single letter (or by a message of length $N = 1$). For instance, Timpson takes this strategy when claims: "*It is crucial to realise that 'information' in Shannon's theory is not associated with individual messages, but rather characterises the source of the messages.*" (Timpson 2004, p. 11). This viewpoint allows him to *define* information in terms of Shannon's theorems; we will come back to this point in Section 6.

An author who incorrectly believes that Shannon's theory cannot deal with the information associated with single states or with individual messages is Fred Dretske (1981). According to him, one of the reasons why Shannon's theory is unable to incorporate semantic content is that semantic notions apply to individual items, while the theory of information is referred to average amounts of information: "*if information theory is to tell us anything about the informational content of signals, it must forsake its concern with averages and tell us something about the information contained in particular messages and signals. For it is only particular messages and signals that have a content*." (Dretske 1981, p. 48). For this reason, he considers necessary to "complete" the theory by defining the individual information $I(s_a)$ as the amount of information generated at the source by the occurrence of a given state s_a, and the individual mutual information $I(s_a;r_a)$ as the information about the occurrence of s_a received at the destination by the occurrence of r_a (*ibid*., p. 52).

Dretske seems to believe that his new definitions, although consistent with the traditional formalism, represent a novelty in the context of Shannon's theory, since they "*are now being assigned a significance, given an interpretation, that they do not have in standard applications of communication theory. They are now being used to define the amount of information associated with particular events and signals*" (Dretske 1981, p. 52). However, this is not the case: when, in Shannon's theory, the entropies of the source and the destination are defined as averages amounts of information, it is clear that the corresponding individual magnitudes can also be defined. When Dretske's proposal was criticized (Timpson 2004, Lombardi 2005), the criticisms did not rely on the fact that it introduces individual amounts of information alien to Shannon's theory, but on a formal mistake in the definition of the individual mutual information $I(s_a;r_a)$, which does not lead to the mutual information $H(S;D)$ when the weighted averages are correctly computed (see Lombardi 2005 for the way in which the error can be amended).

In summary, although it is true that Shannon's theory is not interested in individual amounts of information, this does not mean that those quantities cannot be defined. In fact, in the first paragraph of his paper, Shannon (1948) explicitly says that his own proposal aims at extending the work of Ralph Hartley (1928), where a logarithmic function is introduced as a measure of uncertainty in the case of equiprobability; the Hartley function can be viewed as measuring an individual entropy. In Shannon's theory, where the magnitudes introduced are conceived as average amounts of information, the individual magnitudes must necessarily be part of the formalism, since only in terms of them the averages can be significantly computed. This is precisely the strategy adopted by Shannon in his paper, when he defines an entropy H_i for each state s_i of the source, and defines the entropy of the source "*as the average of these* H_i *weighted in accordance with the probability of occurrence of the states in question*" (Shannon 1948, p. 396).

The discussion about which the most basic notions of Shannon's theory are may seem irrelevant from an instrumental viewpoint, to the extent that the conclusions do not affect the application of the theory. However, it is essential when the interpretation of the concept of information is the issue at stake: in this context is important to decide whereas $I(s_i)$ measures information and $H(S)$ is a weighted average, or the concept of information is directly embodied in $H(S)$, conceived in terms of the coding theorem without relying on the individual information $I(s_i)$. As we will see in Section 9, the epistemic and the physical interpretations of Shannon information adopt the first alternative, whereas the deflationary interpretation takes the second one.

5.- The units of measurement for information

As pointed out in Section 2, the choice of a logarithmic base amounts to a choice of a unit for measuring information. If the base 2 is used, the resulting unit is called '*bit*'. But the natural logarithm can also be used, and in this case the unit of measurement is the *nat,* contraction of *natural unit*. And when the logarithm to base 10 is used, the unit is the *Hartley*. The possibility of using different units to quantify information shows the difference between the amount of information associated with an event and the number of binary symbols necessary to encode it.

For a long time it was quite clear for communication engineers that "bit" was a unit of measurement, and that the fact that a different unit can be used did not affect the very nature of information. However, with the advent of quantum information, the new concept of *qubit* entered the field: a qubit is primarily conceived not as a unit of measurement of quantum information, but as a quantum system of two-states used to encode the information of a source. This is not the appropriate place to analyze this concept and its role in the discussions about quantum information. Nevertheless, it is worth noticing that this way of talking about qubits has gradually seeped into Shannon's theory in the talk about bits. This process led to a progressive reification of the concept of bit, which now is also –and many times primarily– conceived as referring to a classical system of two states. Some authors still distinguish between the two meanings of the concept: "*I would like to distinguish two uses of the word "bit." First, "bit" refers to a unit of information that quantifies the uncertainty of two equiprobable choices. Second, "bit" also refers to a system that can be in one of two discrete states.*" (Duwell 2003, p. 486). But nowadays the conflation between the two meanings is much more frequent: "*The Shannon information $H(X)$ measures in bits (classical two-state systems) the resources required to transmit all the messages that the source produces*." (Timpson 2006, p. 592).

Although very widespread, this undifferentiated use of the term 'bit' sounds odd to the ears of an old communication engineer, for whom the difference between a system and a unit of measurement is deeply internalized. For him, to conflate a bit with a two-state system is like confusing a meter with the Prototype Meter bar, an object made of an alloy of platinum and iridium and stored in the *Bureau International des Poids et Mesures* in Sèvres, near Paris. And saying that the Shannon information $H(X)$ gives a measure "in bits (classical two-state systems)" is like saying that the length L gives a measure "in meters (platinum-iridium bars)".

In order to avoid this kind of confusions about the concept of bit, it might be appropriate to follow the suggestion of Carlton Caves and Christopher Fuchs (1996), who propose to use the term '*cbit*' to name a two-state classical system used to encode Shannon information, by analogy with the two-state quantum system, the qubit, used to encode quantum information (or, at least to encode information by means of quantum resources). This terminology keeps explicit the distinction between the quantity of information produced at the source –an average quantity–, which is usually measured in bits, the alphabet by means of which the messages of the source are encoded, which consist of a number q of symbols, and the systems of q states used to physically implement the code alphabet.

Again, these distinctions may seem an irrelevant matter of detail. Nevertheless, as we will see in the next section, not distinguishing clearly enough between the units of measurement for information and the number of states of the systems used for coding –or the number of symbols of the code alphabet– is a manifestation of the insufficiently clear differentiation between the stage of generation of information and the stage of coding information. And this, in turn, affects the very definition of the concept of information, as will be argued in the next section.

6.- Information and its coding

In the previous section it has been said that the amount of information generated at the source can be expressed in different units of measurement, and this fact is a manifestation of the difference between the amount of information associated with an event and the number of symbols necessary to encode it. Let us consider this claim in more detail.

The source S is a system of n states s_i, which can be thought as the *letters* of an alphabet $A_S = \{s_1, ..., s_n\}$, each with its own probability $p(s_i)$; the sequences of letters (states) are called *messages*. The entropy of the source $H(S)$ can be computed exclusively in terms of these elements –the number of the letters and their probabilities–, and is measured in *bits* when the logarithm to base 2 is used. In turn, the transmitter encodes the messages of the source, and this amounts to performing the conversion between the alphabet of the source, $A_S = \{s_1, ..., s_n\}$, and the code alphabet of the transmitter T, $A_C = \{c_1, ..., c_q\}$, whose q members c_i can be called *symbols*; the sequence of symbols produced by the transmitter and entering the channel is the *signal*. The n-ary source alphabet A_S may be very different depending on the situation, but the code alphabet A_C is more often binary: $q = 2$. In this case, the symbols are *binary digits* (binary alphabet symbols). Finally, the code alphabet A_C can be physically implemented by means of systems of q states; in the particular case that $q = 2$, the two-state systems are *cbits*.

In Shannon's context, coding is a mapping from the source alphabet A_S to the set of finite length strings of symbols from the code alphabet A_C, also called *code-words*. In general, the code-words do not have the same length: the code-word w_i, corresponding to the letter s_i, has a length l_i. This means that coding is a fixed- to variable-length mapping. Therefore, the *average code-word length L* can be defined as:

$$L = \sum_{i=1}^{n} p(s_i)\, l_i \tag{9}$$

L indicates the compactness of the code: the more *L* is low, the more coding is efficient, that is, fewer resources are needed to encode the messages. The Noiseless-Channel Coding Theorem proves that, for very long messages, there is an optimal encoding process such that the average code-word length *L* is as close as desired to the lower bound $L_{\min}$ for *L*:

$$L_{\min} = \frac{H(S)}{\log q} \tag{10}$$

where, when $H(S)$ is measured in bits, log is the logarithm to base 2.

The aim of this formal digression is to emphasize again the difference between the generation of information at the source and the coding of information at the transmitter. The information generated at the source is measured –in average– by $H(S)$, only depends on the features of the source, and can be expressed in bits or in any other unit of measurement. In addition, that information so generated can be encoded by means of a code alphabet of any number of symbols, and the average length of the code-words depends essentially of that number. Only when $H(S)$ is measured in bits and the code alphabet has two symbols (an alphabet of binary digits, $q = 2$), then $\log_2 q = \log_2 2 = 1$, and the noiseless coding theorem establishes the direct relation between the entropy of the source and the lower bound $L_{\min}$ of the average code-word length *L* (see eq. (9)).

In spite of the clear difference between the stages of generation and of coding of information, they are not sufficiently distinguished in the discussions about the meaning of the concept of information. For instance, in the same argumentative context where the two meanings of the word 'bit' are not taken into account (see previous section), Timpson *defines* information in terms of Shannon's theorems: "*the coding theorems that introduced the classical (Shannon, 1948) and quantum (Schumacher, 1995) concepts of information*$_t$ [the technical concept of information] *do not* merely *define measures of these quantities. They also introduce the concept of* what it is *that is transmitted,* what it is *that is measured*." (Timpson 2008, p. 23; emphasis in the original). From this perspective, the meaning of the entropy $H(S)$ is defined by the First Shannon theorem: "*to ask how much information*$_t$ [the technical concept of information] *a source* *produces is ask to what degree is the output of the source* compressible*?*" (Timpson 2008, p. 27; emphasis in the original), and even more clearly, "*The Shannon information $H(X)$* […] *measures how much the messages from the source can be compressed.*" (Timpson 2006, p. 592). Analogously, the meaning of the mutual information $H(S;D)$ is conceived as given by the Second Shannon theorem: "*The most important interpretation of the mutual information does derive from the noisy coding theorem.*" (Timpson 2004, p. 19).

The strategy of defining Shannon information via the coding theorems, although seemingly innocuous, involves several difficulties. First, it turns the theorems into definitions; for instance, the first theorem turns out to be not a theorem but the definition of the entropy $H(S)$ as the average number of bits necessary to encode a letter of the source using an ideal code, and eq. (2) becomes a theorem resulting from a mathematical proof. Second, even accepting the definition of $H(S)$ in terms of the First theorem, what about $H(D)$, which is not covered by Shannon's theorem? If it does not represent information, it is not clear how it can be involved together with $H(S)$ in algebraic operations: how numbers representing different items can be added to or subtracted from each other? If, on the contrary, $H(D)$ is defined by eq. (3), it is not clear why its definition is so different from the definition of $H(S)$, that is, why the symmetry between eqs. (1) and (3) is broken: whereas $H(D)$ is a magnitude that can be defined as an average of the individual amounts of information $I(d_j)$ received at the destination, $H(S)$ is not an average since defined without reference of individual amounts of information.

Moreover, to the extent that entropy is defined in terms of ideal coding and for the case of very long messages (strictly speaking, for $N \to \infty$), what happens in the case of non-ideal coding?, can we still say that a same amount of information can be better or worse encoded? Can short binary messages be conceived as embodying information even if not covered by the coding theorems? Furthermore, although in general the transmitter encodes the message and produces a signal suitable for transmission over the channel, in certain cases of communication the message is not encoded; for instance, in traditional telephony the transmitter is a mere transducer that changes sound pressure into an electrical current. If information is defined as a measure of the compressibility of messages during coding, how to talk about information in those situations where no coding is involved?

The arguments of this and the two previous sections show up the need to be clear about the different elements in Shannon's theory. As we will see in Section 9, conflating in principle distinct concepts, although may not affect the everyday applications of the theory –in which information is usually measured in bits and codified with binary digits–, affects the interpretation of the concept of information.

7.- The relative nature of information

As stressed in Section 3, in Shannon's theory it makes sense to talk about the information associated with the occurrence of individual states. However, this does not mean that Shannon information can be defined independently of the consideration of the systems involved in the communication arrangement as systems that produce states with their corresponding probabilities. And the characterization of those systems as sources or destinations is not unique, but depends on the particular case of interest. In other words, information is not an absolute magnitude, but is *relative* to the whole communication situation.

The relative nature of information is usually stressed mainly by those who link information with knowledge. From this perspective, information depends on the knowledge about the source available at the destination before the transmission: "*the datum point of information is then the whole body of knowledge possessed at the receiving end before the communication*" (Bell 1957, p. 7). But the different ways of characterizing the source may depend not only on epistemic reasons, but also on pragmatic matters, such as the particular interest that underlies the definition of the whole communication arrangement as such. For instance, a roulette wheel can be described as a source with 37 states when we are interested in a single number, or as a source with 3 states when we are interested in color: although the physical system is the same, in informational terms the two sources are completely different. The choice between one alternative or the other depends exclusively on pragmatic reasons.

But not only is the characterization of the source of information relative. As noticed in Section 3, the success criterion in Shannon's theory is completely arbitrary. The criterion is given by a one-to-one mapping from the source alphabet to the destination alphabet that is not unique or essentially fixed: on the contrary, it is also usually determined by pragmatic reasons. As Duwell points out when considering the success of communication as a convention: "*one might simply treat real information sources as producing an abstract sequence as on the bare Shannon theory, and have a success condition be relative to an arbitrarily chosen one-one function between the information source and destination*." (Duwell 2008, p. 201).

It is interesting to notice that the relativity of information is usually not thematized in the textbooks about Shannon's theory. Perhaps the reason for this is an implicit identification between relativity and subjectivity: it is supposed that admitting the relative character of information would threaten the scientific status of Shannon's theory. But, of course, this conclusion is drawn from an incorrect identification. The pragmatic decisions that underlie the characterization of the communication arrangement must not be conceived as a subjective ingredient, but as a reference frame with respect to which the magnitudes are defined without losing their objectivity. The relativization of objective magnitudes is very frequent in sciences; in this sense, information is not different from velocity and simultaneity, which are relative to a certain reference frame, but not for this reason are less objective. In fact, only on the basis of a conception of information as an objective and quantifiable magnitude, a formal and precise theory with many technological applications could be formulated.

8.- The theoretical neutrality of Shannon information

For many years after the publication of the 1948 paper, the theory was successfully applied to technological problems and nobody was interested in discussing what physical theory, if any, underlay Shannon's proposal. However, during the last decades we are witnessing the explosion of a new field, that of the so-called quantum information. At present it is considered that the work of Benjamin Schumacher (1995) offers the quantum analog of Shannon's theory, and that in the new formalism the von Neumann entropy measures quantum information, playing a role analogous to that of the Shannon information in Shannon's theory. So, since the emergence of quantum information, the idea that Shannon information is "classical" and essentially different from quantum information has progressively permeated the information community.

A clear example of this position is given by Časlav Bruckner and Anton Zeilinger (2001), who argue that the Shannon information is closely linked to classical concepts, in particular, to the classical conception of measurement and, for this reason, it is not appropriate as a measure of information in the quantum context. Besides of an incorrect interpretation of the so-called grouping axiom (Shannon 1948, p. 393; for a demolishing criticism of Brukner and Zeilinger's argument, see Timpson 2003), the authors claim that the concept has no operational meaning in the quantum case because quantum observables have no definite values pre-existing to measurements: "*The nonexistence of well-defined bit values prior to and independent of* *observation suggests that the Shannon measure, as defined by the number of binary questions needed to determine the particular* observed *sequence 0's and 1's, becomes problematic and even untenable in defining our uncertainty as given* before *the measurements are performed*." (Brukner and Zeilinger 2001, p. 1; emphasis in the original). But, as Timpson (2003) clearly argues, there is nothing in Shannon's theory that requires actual sequences of states-letters in the source S to define the entropy $H(S)$. In fact, as shown by the presentation of the formalism of the theory in the previous sections, $H(S)$ depends on the statistical features of the source, that is, its states and the corresponding probabilities. But nothing is said about how the probabilities are determined nor about their interpretation: they may be conceived as propensities theoretically computed, or as frequencies previously measured; in any case, once the system S turns out to play the role of source of information, the measurement of an actual sequence of states is not necessary to define and compute the entropy $H(S)$.

From a more general perspective, the central point to emphasize is that the definition of the elements involved in Shannon's theory is independent of their physical substratum: the states-letters of the source are not physical states but are *implemented* by physical states, which may be of very varied nature. And the same can be said about the channel, which embodies the correlations between source and destination: it does not matter how those correlations are established and physically "materialized"; what only matters is that they link the states of the source and the states of the destination (we will come back to this point in the next section). This means that Shannon's theory is not "classical" in any meaningful physical sense of the term 'classical', and "*can be applied to any communication system regardless whether its parts are best described by classical mechanics, classical electrodynamics, quantum theory, or any other physical theory*." (Duwell 2003, p. 480).

Once it is acknowledged that Shannon's theory is neutral with respect to the physical theory on the basis of which the communication arrangement is implemented, it is easy to see that there is no obstacle to its application to the quantum context. In fact, when the letters of a source S with entropy $H(S)$ are encoded by means of orthogonal quantum states but decoded in a different basis, there is a loss of information that can be represented in Shannon's terms as an equivocation E (see eq. (7)), such that the mutual information (the average amount of information generated at the source S and received at the destination D, see eq. (5)) is computed as de difference between the entropy of the source and the loss represented by E: $H(S;D) = H(S) - E$ (see Schumacher 1995, p. 2739).

Up to this point we have shown that Shannon's theory is a quantitative theory whose elements have no semantic dimension, and that it defines average amounts of information that can be measured in different units of measurement and whose values are relative to the whole communication arrangement. Moreover, Shannon's theory is not tied to a particular physical theory, but is independent of its physical implementation. If Shannon information is what is thematized by Shannon's theory, the agreement about all these features of the theory might suggest that there is a clear interpretation of the concept of Shannon information shared by the whole information community. But this is not the case at all: the concept of Shannon information is still a focus of much debate.

9.- Interpreting the concept of information

As stressed in the previous section, Shannon's theory is theoretically neutral regarding physics. Perhaps this feature is what leads some authors to consider that Shannon information is not physical. An active representative of this position is Timpson, for whom the dictum '*Information is physical*', applied to the technical concept of information, if not trivial –meaning that some physically defined quantity is physical–, is false precisely because 'information' is an abstract noun (Timpson 2004, 2008, 2013). Since 'information' is an abstract noun, "*it doesn't serve to refer to a material thing or substance*." (Timpson 2004, p. 20). This "deflationary" view of information, according to which "*there is not a question* […] *of 'the information' being a referring term*." (Timpson 2006, p. 599), is in line with his interpretation of the entropy $H(S)$, not as quantifying something produced by the source, but as a measure of compressibility of messages.

Independently of the reasons to reach the conclusion about the abstract nature of information (recall our discussion in Section 3), it is worth noting that the fact that an item is abstract does not imply that it is not physical, and even less that its name does not refer. It is not necessary to be a substance, or a concrete thing, or a material entity, to be physical. The realm of physics is populated by countless properties, usually referred to as 'observables', which are not substances nor concrete or material things; only from an extreme nominalist perspective the existence of physical properties can be called into question. Moreover, physics, in its evolution, tends to perform a substantialization of certain concepts: from originally being conceived as properties, certain magnitudes turn into substances, not in the sense of becoming kinds of stuff, referents of mass nouns, but in the Aristotelian sense of being objects of predication but not predicable of anything else. The paradigmatic case is that of energy, which also seems to be something non-material but, at the same time, is one of the fundamental physical concepts and plays a central unifying role in physics. Summing up, the acknowledgement of the abstract nature of information does not release us from the philosophical inquiry about the interpretation of the concept of information.

The concept most usually connected with the notion of information is that of knowledge: information provides knowledge, modifies the state of knowledge of those who receive it. It can be supposed that the link between information and knowledge is a feature of the everyday notion of information and not of Shannon's concept (see Timpson 2004, 2013), but the literature on the subject shows that this is not the case: that link can be frequently found both in philosophy and in science. For instance, taking Shannon's theory as the underlying formalism for his proposal, Fred Dretske says: "*information is a commodity that, given the right recipient, is capable of yielding knowledge.*" (1981, p. 47). Some authors devoted to special sciences are persuaded that the core meaning of the concept of information, even in its technical sense, is linked to the concept of knowledge. In this trend, Jon M. Dunn defines information as "*what is left of knowledge when one takes away believe, justification and truth*" (2001, p. 423). Also physicists frequently speak about what we know or may know when dealing with information. For instance, Zeilinger even equates information and knowledge when he says that "[w]*e have knowledge, i.e., information, of an object only through observation*" (1999, p. 633) or, with Bruckner, "[f]*or convenience we will use here not a measure of information or knowledge, but rather its opposite, a measure of uncertainty or entropy.*" (2009, pp. 681-682). In a traditional textbook about Shannon's theory applied to engineering it can also be read that information "*is measured as a difference between the state of knowledge of the recipient before and after the communication of information.*" (Bell 1957, p. 7). Although not regarding Shannon's theory but in the quantum context, Christopher Fuchs adheres to Bayesianism regarding probabilities and, as a consequence, advocates for an epistemic interpretation of information (see Caves, Fuchs and Schack 2002).

Although from the epistemic perspective information is not a physical item, in general it is assumed that the possibility of acquiring knowledge about the source of information by consulting the state of the destination is rooted in the nomic connection between them, that is, in the lawfulness of the regularities underlying the whole situation: "*The conditional probabilities used to compute noise, equivocation, and amount of transmitted information* […] *are all determined by the lawful relations that exist between source and signal. Correlations are irrelevant unless these correlations are a symptom of lawful connections*" (Dretske 1981, p. 77). In fact, if those conditional probabilities represented accidental, merely *de facto* correlations, the states in the destination would tell us nothing about the state of the source. Nevertheless, this appeal to lawful connections opens new questions for the epistemic view. Noise and equivocation are indeed defined in terms of nomic correlations, but in what sense they supply knowledge? Whereas the mutual information $H(S;D)$ can be easily interpreted as a measure of the knowledge about the source obtained at the destination, noise and equivocation do not measure knowledge but, on the contrary, are obstacles to knowledge acquisition. It is not easy to see how noise, which can be generated outside of the communication arrangement and has no relation with the source of information (think, for instance, in white noise in a radio receiver), can be conceived as something carrying or yielding knowledge. A way out of this problem might be to suppose that only the entropies of source and destination and the mutual information, but not noise and equivocation, can be meaningfully conceptualized as measures of knowledge. But this answer would lead to admit the possibility of adding and subtracting variables referring to different kinds of items, in this case knowledge and something different from knowledge (see, e.g., eq. (5)), a practice absolutely not allowed in mathematized sciences.

A different view about information is that which considers information as a physical magnitude. This is the position of many physicists (see, *e.g.*, Rovelli 1996) and most engineers, for whom the essential feature of information consists in its capacity to be generated at one point of the physical space and transmitted to another point; it can also be accumulated, stored and converted from one form to another. From this perspective, the link with knowledge is not a central issue, since the transmission of information can be used only for control purposes, such as operating a device at the destination end by modifying the state of the source.

In general, the physical interpretation of information appears strongly linked with the idea expressed by the well-known *dictum* 'no information without representation': the transmission of information between two points of the physical space necessarily requires an information-bearing signal, that is, a physical process propagating from one point to the other. Rolf Landauer is an explicit defender of this position when he claims that "[i]*nformation is not a disembodied abstract entity; it is always tied to a physical representation. It is represented by engraving on a stone tablet, a spin, a charge, a hole in a punched card, a mark on a paper, or some other equivalent*." (1996, p. 188; see also Landauer 1991). This view is also adopted by some philosophers of science; for instance, Peter Kosso states that "*information is transferred between states through interaction.*" (1989, p. 37). The need of a carrier signal sounds natural in the light of the generic idea that physical influences can only be transferred through interactions. On this basis, information is conceived by many physicists as a physical entity with the same ontological status as energy; it has also been claimed that its essential property is the power to manifest itself as structure when added to matter (Stonier 1990, 1996).

As we argued above, pace Timpson, the abstract nature of information is not an obstacle to a physical interpretation of the concept. Nevertheless, it might be thought that the neutrality of Shannon information with respect to the physical theory –or theories– involved in the implementation of the communication arrangement is really a trouble for the physical interpretation. However, this conclusion is not unavoidable. In fact, many physical concepts, through the evolution of the discipline, have experienced a process of abstraction and generalization in such a way that, at present, they are no longer tied to a specific theory but permeate the whole of physics. Again, energy is the most conspicuous example: since essentially present in all the theories of physics, it is not tied to one in particular; it has different physical manifestations in different domains; nevertheless, the concept of energy is perhaps the physical concept par excellence. Mutatis mutandis, the same can be said of information, and therefore its theoretical neutrality is not an insurmountable obstacle to interpret the concept in physical terms.

The difference between the epistemic and the physical interpretations of information is not merely nominal, but may yield different conclusions regarding certain common physical situations. For instance, in the influential philosophical tradition that explains scientific observation in terms of information (Shapere 1982, Brown 1987, Kosso 1989), the way in which information is conceived leads to very different consequences regarding observation. This turns out to be particularly clear in the so-called 'negative experiments' (see Jammer 1974), in which it is assumed that an object or event has been observed by noting the absence of some other object or event. From the informational view of scientific observation, observation without a direct physical interaction between the observed object and an appropriate destination is only admissible from an epistemic interpretation of information. According to a physical interpretation, by contrast, without interaction there is no observation: the presence of the object is only inferred (see Lombardi 2004).

Let us consider a source S that transmits information to two physically isolated receivers R_A and R_B via a certain physical link. In this case, the correlations between the states of the two receivers are not accidental, but they result from the physical dependence of the states of R_A and R_B on the states of S. Nevertheless, there is no physical interaction between the two receivers. Also in this case the informational description of the situation is completely different from the viewpoints given by the two interpretations of the concept of information. According to the physical interpretation, it is clear that there is no information transmission between R_A and R_B to the extent that there is no physical signal between them. However, from an epistemic interpretation, nothing prevents us from admitting the existence of an *informational* link between the two receivers. In fact, we can define a communication channel between R_A and R_B because it is possible to learn something about R_B by looking at R_A and vice versa: "*from a theoretical point of view* [. . .] *the communication channel may be thought of as simply the set of depending relations between* [a system] *S and* [a system] *R. If the statistical relations defining equivocation and noise between S and R are appropriate, then there is a channel between these two points, and information passes between them, even if there is no direct physical link joining S with R.*" (Dretske 1981, p. 38). The receiver R_B may even be farther from the source S than R_A, so that the events at R_B may occur later than those at R_A. Nevertheless, this is irrelevant from the epistemic view of information: despite the fact that the events at R_B occur later, R_A carries information about what will happen at R_B.

Although the description of the situation given from the epistemic view is completely consistent, there is still something in it that sounds odd. In fact, communication implies that, at some place, someone does something that has consequences somewhere else. But in the case of the two receivers, nothing can be done, say, at the R_A end that will affect what happens in the R_B end. In other words, the change of the state of R_A cannot be used to control the state of R_B; so, something of the usual conception of the process of transmitting information is missing. The example of the two receivers would be analogous to the case of the EPR-type experiments, characterized by theoretically well-founded correlations between two spatially separated particles. During many years it was repeated that information cannot be sent between both particles because the propagation of a superluminal signal from one particle to the other is impossible: there is no information-bearing signal that can be modified at one point of space in order to carry information to the other spatially separated point. For the defender of the physical interpretation of information these arguments act as a silver bullet for the epistemic view, since they make clear the need of a physical carrier of information between source and destination; it is this physical signal that allows us to say that what happens at the source causes what happens at the destination.

However, things are not so easy when quantum mechanics comes into play with the case of teleportation. Broadly speaking, an unknown quantum state is transferred from Alice to Bob with the assistance of a shared pair prepared in an entangled state and of two classical bits sent from Alice to Bob (the description of the protocol can be found in any textbook on the matter). In general, the idea is that the very large (strictly infinite) amount of information required to specify the teleported state is transferred from Alice to Bob by sending only two bits. When addressing this problem, many physicists try to find a physical link between Alice and Bob that could play the role of the carrier of information. For instance, Roger Penrose (1998) and Richard Jozsa (1998, 2004) claim that information may travel backwards in time: "*How is it that the continuous 'information' of the spin direction of the state that she wishes to transmit* […] *can be transmitted to Bob when she actually sends him only two bits of discrete information? The only other link between Alice and Bob is the quantum link that the entangled pair provides. In spacetime terms this link extends back into the past from Alice to the event at which the entangled pair was produced, and then it extends forward into the future to the event where Bob performs his*." (Penrose 1998, p. 1928). According to David Deutsch and Patrick Hayden (2000), the information travels hidden in the classical bits. These physicists do not explicitly acknowledge that the problem derives from the physical interpretation of information to which they are strongly tied, and that an epistemic view would not commit them to find a physical channel between Alice and Bob. Perhaps for this reason Timpson prefers to directly reject the physical interpretation and designs his deflationary view to support such rejection.

The case of teleportation shows that, although the mere correlation is not sufficient for communication of information, asking for a physical signal acting as a carrier of information from source to destination is a too strong requirement that leads to artificial solutions as those of backwards flowing information or of classically hidden information. What a non-epistemic interpretation of information needs is the idea that what happens at the source causes what happens at the destination, but with a concept of causality that does not rely on physical interactions or space-time lines connecting the states of the source with the states of the destination: causality cannot be conceived in terms of energy flow (Fair 1979, Castañeda 1984), physical processes (Russell 1948, Dowe 1992), or property transference (Ehring 1997, Kistler 1998). Perhaps good candidates for conceptualizing the informational links from a non-epistemic stance are the manipulability theories of causation, according to which causes are to be regarded as devices for manipulating effects (Price 1991, Menzies and Price 1993, Woodward 2003). The rough idea is that, if C is genuinely a cause of E, then if one can manipulate C in the right way, this should be a way of manipulating or changing E (for an introduction, and also criticisms, see Woodward 2013). The view of causation as manipulability is widespread among statisticians, theorists of experimental design and many social and natural scientists, as well as in causal modeling. In the present context we are not interested in discussing whether this is the correct or the best theory of causation in general, or whether it can account for all the possible situations usually conceived as causation. Here it suffices to notice that the manipulability view may be particularly useful to elucidate the concept of Shannon information, in the context of a theory for which "[t]*he fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.*" (Shannon 1948, p. 379). This view blocks situations like those of the two correlated receivers as cases of information transmission; but, at the same time, it admits cases, such as teleportation, in which there is a certain control of what happens in the destination end by means of actions at the source end, in spite of the absence of any physical signal between the two ends of the communication arrangement.

10.- A formal view of Shannon information

In the traditional textbooks about information, Shannon's theory is usually introduced from a physical perspective, although frequently including epistemic elements, without realizing the difference between the two interpretations. However, this has changed in the last decades. As a result, at present the textbooks on the matter begin by treating information in an exclusively formal way. There are no sources, destinations or signals; the basic concepts are introduced in terms of random variables and probability distributions over their possible values. The traditional case of communication is introduced only after the formal presentation, as one of the many applications of the theory. This formal view of information already appears in the classical books of Aleksandr Khinchin (1957) and Fazlollah Reza (1961), who conceive information theory as a new chapter of the mathematical theory of probability. But perhaps the best-known example of this approach is the presentation offered by Thomas Cover and Joy Thomas in his book *Elements of Information Theory* (1991), who clearly explain their viewpoint just from the beginning: "*Information theory answers two fundamental questions in communication theory: what is the ultimate data compression* [...] *and what is the ultimate transmission rate of communication* [...]. *For this reason some consider information theory to be a subset of communication theory. We will argue that it is much more. Indeed, it has fundamental contributions to make in statistical physics (thermodynamics), computer sciences (Kolmogorov complexity or algorithmic complexity), statistical inference (Occam's Razor: 'The simplest explanation is best') and to probability and statistics (error rates for optimal hypothesis testing and estimation)*" (Cover and Thomas 1991, p. 1).

From this perspective, the first step is to define two discrete random variables with alphabets A and B, and probability mass functions $p(x) = \Pr(X = x)$, with $x \in A$, and $p(y) = \Pr(Y = y)$, with $y \in B$, respectively (the presentation can be extrapolated to continuous variables). On this basis, the entropy of the variables X and Y, $H(X)$ and $H(Y)$ are defined as usual. Other relevant magnitudes are the joint entropy $H(X,Y)$ of the variables X and Y, computed in terms of the joint distribution $p(x, y)$, and the conditional entropies $H(X/Y)$ of X given Y and $H(Y/X)$ of Y given X, computed in terms of the conditional probabilities $p(x/y)$ and $p(y/x)$, respectively. In turn, the mutual entropy $H(X;Y)$ is defined as the relative entropy between the joint distribution $p(x, y)$ and the product distribution $p(x)p(y)$. Since the presentation is merely formal, all the above definitions can be extended to the case of more than two random variables, leading, for instance, to the entropy $H(X_1, X_2,..., X_n)$ of a collection random variables, or to the conditional mutual entropy $H(X_1, X_2,..., X_n/Y)$ of the random variables $X_1, X_2,..., X_n$ given Y.

Of course, in this formal context information has nothing to do with physical theories or propagation of signals. But this view also cuts any link between information and knowledge to the extent that it does not require an underlying network of lawful relations: the probabilities can be computed on the basis of merely *de facto* frequencies and correlations. As we have stressed above, when the correlation between two variables is merely accidental, the value of one of them tells us nothing about the value of the other. Therefore, from this formal approach the basic intuition according to which information modifies the state of knowledge of those who receive such information gets lost.

The defender of the epistemic interpretation of information may consider that this is a too high price to pay to retain a formally precise formulation of information theory. However, the formal view has its own advantage: by turning information into a formal concept, it makes the theory applicable to a variety of fields. Communication by means of physical signals is only one among those fields, as well as the entanglement assisted communication supporting teleportation.

In more precise terms, the formal view endows the concept of information with a generality that makes it a powerful formal tool for science. This means that the word ‘information’ does not belong to the language of factual sciences or to ordinary language: it has no semantic content. It is not only that messages have no semantic content, but that the concept of information is a purely mathematical concept, whose “meaning” has only a syntactic dimension. It is precisely from its syntactic nature that the generality of the concept derives.

From this formal perspective, the relationship between the word ‘information’ and the different views of information is the logical relationship between a mathematical object and its interpretations, each one of which endows the term with a specific referential content. The epistemic view, then, is only one of the many different interpretations, which may be applied in psychology and in cognitive sciences: the concept of information can be used to conceptualize the human abilities of acquiring knowledge (see e.g. Hoel, Albantakis and Tononi 2013). The epistemic interpretation might also serve as a basis for the philosophically motivated attempts to add a semantic dimension to a formal theory of information (MacKay 1969, Nauta 1972, Dretske 1981)

In turn, the physical view, which conceives information as a physical magnitude, is clearly the appropriate interpretation in communication theory, where the main problem consists in optimizing the transmission of information by means of physical means, whether carrier signals, whose energy and bandwidth is constrained by technological and economic limitations, or quantum entanglement, which must be protected from decoherence. But this is not the only possible physical interpretation: when S is interpreted as a macrostate compatible with many equiprobable microstates, $I(S)$ represents the Boltzmann thermodynamic entropy of S. Furthermore, in computer sciences an algorithmic information may be defined, such that, if S is interpreted as a binary string of finite length, $I(S)$ can be related with the algorithmic complexity of *S*. There are also non-traditional applications, as those based on the relation between Shannon entropy and gambling (see, e.g., Cover and Thomas 1991, Chapter 6) or between Shannon entropy and investment in stock market (see, e.g., Cover and Thomas 1991, Chapter 15).

Summing up, maybe it is time to set aside the monistic stances about information, and to adopt a pluralist position, according to which the different views are no longer rival, but different interpretations of a single formal concept. Each one of these interpretations is legitimate to the extent that its application is useful in a certain scientific or technological field. In this sense, the formal view is in resonance not only with the wide and strong presence of the concept of information in all contemporary human activities, but also with Shannon's position when claiming: "*The word 'information' has been given different meanings by various writers in the general field of information theory*. [...] *It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field*." (Shannon 1993, 180).

11.- Conclusions

Despite of its formal precision and its great many applications, Shannon's theory still offers an active terrain of debate when the interpretation of its main concepts is the task at issue. In this article we have tried to analyze certain points that still remain obscure or matter of discussion, and whose elucidation contribute to the assessment of the different interpretative proposals about the concept of information. Moreover, the present argumentation might shed light on the problems related with the so-called 'quantum information theory', in particular as formulated by Benjamin Schumacher (1995) on the basis of an explicit analogy with the first Shannon coding theorem. Furthermore, the discussion about the interpretation of the concepts involved in Shannon's theory would turn out to be particularly relevant if, as some believe, there were not two kinds of information −classical and quantum−, but only information encoded in different ways.

12.- References

Adriaans, P. (2013). "Information." In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), URL = <http://plato.stanford.edu/archives/fall2013/entries/information/>.

Adriaans, P. and van Benthem, J. (eds) (2008). *Handbook of Philosophy of Information*. Amsterdam: Elsevier Science Publishers.

Bar-Hillel, Y. (1964). *Language and Information: Selected Essays on Their Theory and Application*. Reading, MA: Addison-Wesley.

Bar-Hillel, Y. and Carnap, R. (1953). "Semantic Information." *The British Journal for the Philosophy of Science*, **4**: 147-157.

Barwise, J. and Seligman, J. (1997). *Information Flow. The Logic of Distributed Systems*. Cambridge Cambridge University Press.

Bell, D. (1957). *Information Theory and its Engineering Applications*. London: Pitman & Sons.

Brown, H. (1987). *Observation and Objectivity*. Oxford: Oxford University Press.

Brukner, Č. and Zeilinger, A. (2001). "Conceptual Inadequacy of the Shannon Information in Quantum Measurements." *Physical Review A*, **63**: #022113.

Brukner, Č. and Zeilinger, A. (2009). "Information Invariance and Quantum Probabilities." *Foundations of Physics*, **39**: 677-689.

Castañeda, H.-N. (1984). "Causes, Causity, and Energy." Pp. 17-27, in P. French, T. Uehling, Jr., and H. Wettstein (eds.), *Midwest Studies in Philosophy IX*. Minneapolis: University of Minnesota Press.

Caves, C. M. and Fuchs, C. A. (1996). "Quantum Information: How Much Information in a State Vector?" In A. Mann and M. Revzen (eds.), *The* Dilemma *of Einstein, Podolsky and Rosen - 60 Years Later*. *Annals of the Israel Physical Society*, **12**: 226-257 (see also quant-ph/9601025).

Caves, C. M., Fuchs, C. A., and Schack, R. (2002b). "Unknown Quantum States: The Quantum de Finetti Representation." *Journal of Mathematical Physics*, **43**: 4537-4559.

Chaitin, G. (1987). *Algorithmic Information Theory*. New York: Cambridge University Press.

Cover, T. and Thomas, J. A. (1991). *Elements of Information Theory*. New York: John Wiley & Sons.

Deutsch, D. and Hayden, P. (2000). "Information Flow in Entangled Quantum Systems." *Proceedings of the Royal Society of London A*, **456**: 1759-1774.

Dowe, P. (1992). "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory." *Philosophy of Science*, **59**: 195-216.

Dretske, F. (1981). *Knowledge & the Flow of Information*. Cambridge MA: MIT Press.

Dunn, J. M. (2001). "The Concept of Information and the Development of Modern Logic." Pp. 423-427, in W. Stelzner (ed.), *Non-classical Approaches in the Transition from Traditional to Modern Logic*. Berlin: de Gruyter.

Duwell, A. (2003). "Quantum Information Does Not Exist." *Studies in History and Philosophy of Modern Physics*, **34**: 479-499.

Duwell, A. (2008). "Quantum Information Does Exist." *Studies in History and Philosophy of Modern Physics*, **39**: 195-216.

Ehring, D. (1986). "The Transference Theory of Causality." *Synthese*, **67**: 249-58

Fair, D. (1979). "Causation and the Flow of Energy." *Erkenntnis*, **14**: 219-250.

Fisher, R. (1925). "Theory of Statistical Estimation." *Proceedings of the Cambridge Philosophical Society*, **22**: 700-725.

Floridi, L. (2010). *Information – A Very Short Introduction*. Oxford: Oxford University Press.

Floridi, L. (2011). *The Philosophy of Information*, Oxford; Oxford University Press.

Floridi, L. (2013). "Semantic Conceptions of Information." In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition), URL = <http://plato.stanford.edu/archives/spr2013/entries/information-semantic/>.

Hartley, R. (1928). "Transmission of Information." *Bell System Technical Journal*, **7**: 535-563.

Hoel, E., Albantakis, L. y Tononi, G. (2013). "Quantifying Causal Emergence Shows that Macro Can Beat Micro." *Proceedings of the National Academy of Sciences*, **110**: 19790-19795.

Jammer, M. (1974). *The Philosophy of Quantum Mechanics*. New York: John Wiley & Sons.

Jozsa, R. (1998). "Entanglement and Quantum Computation." Pp. 369-79, in S. Huggett, L. Mason, K. P. Tod, S. T. Tsou, and N. M. J. Woodhouse (eds.), *The Geometric Universe*. Oxford: Oxford University Press.

Jozsa, R. (2004). "Illustrating the Concept of Quantum Information." *IBM Journal of Research and Development*, **4**: 79-85.

Khinchin, A. (1957). *Mathematical Foundations of Information Theory*. New York: Dover.

Kistler, M. (1998). "Reducing Causality to Transmission." *Erkenntnis*, **48**: 1-24.

Kosso, P. (1989). *Observability and Observation in Physical Science*. Dordrecht: Kluwer.

Landauer, R. (1991). "Information is Physical." *Physics Today*, **44**: 23-29.

Landauer, R. (1996). "The Physical Nature of Information." *Physics Letters A*, **217**: 188-193.

Lombardi, O. (2004). "What is Information?" *Foundations of Science*, **9**: 105-134.

Lombardi, O. (2005). "Dretske, Shannon's Theory and the Interpretation of Information." *Synthese*, **144**: 23-39.

MacKay, D. (1969). *Information, Mechanism and Meaning*. Cambridge MA: MIT Press.

Menzies, P. and Price, H. (1993). "Causation as a Secondary Quality." *British Journal for the Philosophy of Science*, **44**: 187-203.

Nauta, D. (1972). *The Meaning of Information*. The Hague: Mouton.

Penrose, R. (1998). "Quantum Computation, Entanglement and State Reduction." *Philosophical Transactions of the Royal Society of London A*, **356**: 1927-1939.

Price, H. (1991). "Agency and Probabilistic Causality." *British Journal for the Philosophy of Science*, **42**: 157-76.

Reza, F. (1961). *Introduction to Information Theory*. New York: McGraw-Hill.

Rovelli, C. (1996). "Relational Quantum Mechanics." *International Journal of Theoretical Physics*, **35**: 1637-1678.

Russell, B. (1948). *Human Knowledge: Its Scope and Limits*. New York: Simon and Schuster.

Schumacher, B. (1995). "Quantum Coding." *Physical Review A*, **51**: 2738-2747.

Shannon, C. (1948). "The Mathematical Theory of Communication." *Bell System Technical Journal*, **27**: 379-423.

Shannon, C. (1993). *Collected Papers*, N. Sloane and A. Wyner (eds.). New York: IEEE Press.

Shannon, C. and Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana and Chicago: University of Illinois Press.

Shapere, D. (1982). “The Concept of Observation in Science and Philosophy.” *Philosophy of Science*, **49**: 485-525.

Stonier, T. (1990). *Information and the Internal Structure of the Universe: An Exploration into Information Physics*. New York-London: Springer.

Stonier, T. (1996). “Information as a Basic Property of the Universe.” *Biosystems*, **38**: 135-140.

Timpson, C. (2003). “On a Supposed Conceptual Inadequacy of the Shannon Information in Quantum Mechanics.” *Studies in History and Philosophy of Modern Physics*, **34**: 441-68.

Timpson, C. (2004). *Quantum Information Theory and the Foundations of Quantum Mechanics*. PhD diss., University of Oxford (quant-ph/0412063).

Timpson, C. (2006). “The Grammar of Teleportation.” *The British Journal for the Philosophy of Science*, **57**: 587-621.

Timpson, C. (2008). “Philosophical Aspects of Quantum Information Theory.” Pp. 197-261, in Dean Rickles (ed.), *The Ashgate Companion to the New Philosophy of Physics*. Aldershot: Ashgate Publishing. Page numbers are taken from the on.line version: arXiv:quant-ph/0611187.

Timpson, C. (2013). *Quantum Information Theory and the Foundations of Quantum Mechanics*. Oxford: Oxford University Press.

Tribus, M. and McIrving, E. C. (1971). “Energy and Information.” *Scientific American*, **225**: 179-188.

Wetzel, L. (2011). “Types and Tokens.” In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), URL = <http://plato.stanford.edu/archives/spr2014/entries/types-tokens/>.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, J. (2013). “Causation and Manipulability.” In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), URL = <http://plato.stanford.edu/archives/win2013/entries/causation-mani/>.

Zeilinger, A. (1999). “A Foundational Principle for Quantum Mechanics.” *Foundations of Physics*, **29**: 631-643.